

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日

Date of Application:

2003年 1月 9日

出 願 番 号

Application Number:

特願2003-002978

[ST.10/C]:

[JP 2003-002978]

出 願 人

Applicant(s):

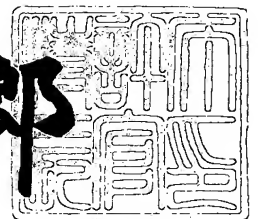
沖電気工業株式会社

Inventor: Yasuhiro KAWAKITA
Atty Dkt: 31759-190542
Customer No: 26644

2003年 4月 1日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3022446

【書類名】 特許願

【整理番号】 KN002568

【提出日】 平成15年 1月 9日

【あて先】 特許庁長官 太田 信一郎 殿

【国際特許分類】 G06F 15/00

【発明者】

 【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

 【氏名】 川北 泰広

【発明者】

 【住所又は居所】 東京都港区虎ノ門1丁目7番12号 沖電気工業株式会
社内

 【氏名】 池野 篤司

【特許出願人】

 【識別番号】 000000295

 【氏名又は名称】 沖電気工業株式会社

 【代表者】 篠塚 勝正

【代理人】

 【識別番号】 100090620

 【弁理士】

 【氏名又は名称】 工藤 宣幸

【先の出願に基づく優先権主張】

 【出願番号】 特願2002-187695

 【出願日】 平成14年 6月27日

【手数料の表示】

 【予納台帳番号】 013664

 【納付金額】 21,000円

【提出物件の目録】

 【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【包括委任状番号】 9006358

【プルーフの要否】 要

【書類名】 明細書

【発明の名称】 電子文書情報拡充装置、方法及びプログラム、並びに、電子文書情報拡充プログラムを記録した記録媒体

【特許請求の範囲】

【請求項 1】 電子文書に対し、それが有しない要素の情報を拡充する電子文書情報拡充装置であって、

入力された電子文書から、それに含まれているデータの所在情報を抽出する所在情報抽出手段と、

抽出された所在情報に基づき、補充し得る外部データを取得する外部データ取得手段と、

取得した外部データから、入力された上記電子文書で不足する要素の情報を拡充する情報補完手段と

を備えることを特徴とする電子文書情報拡充装置。

【請求項 2】 入力された上記電子文書を解析して情報単位に分割する情報解析手段をさらに備え、

上記所在情報抽出手段、上記外部データ取得手段及び上記情報補完手段が、分割された上記各情報単位に対して所定の処理を行う

ことを特徴とする請求項 1 に記載の電子文書情報拡充装置。

【請求項 3】 上記外部データ取得手段は、抽出された所在情報に基づき、外部データを取得する際、外部データと、情報拡充対象の電子文書又は情報単位の内容とが類似していることを判定した上で取得することを特徴とする請求項 1 又は 2 に記載の電子文書情報拡充装置。

【請求項 4】 上記外部データ取得手段は、抽出された所在情報が示す場所の外部データと、その外部データが含む所在情報をたどって到達し得る外部データとを取得可能な外部データとしていることを特徴とする請求項 1 ～ 3 のいずれかに記載の電子文書情報拡充装置。

【請求項 5】 電子文書に対し、それが有しない要素の情報を拡充する電子文書情報拡充方法であって、

入力された電子文書から、それに含まれているデータの所在情報を抽出する所在情報抽出工程と、

抽出された所在情報に基づき、補充し得る外部データを取得する外部データ取得工程と、

取得した外部データから、入力された上記電子文書で不足する要素の情報を拡充する情報補完工程と

を有することを特徴とする電子文書情報拡充方法。

【請求項 6】 入力された上記電子文書を解析して情報単位に分割する情報解析工程をさらに有し、

上記所在情報抽出工程、上記外部データ取得工程及び上記情報補完工程が、分割された上記各情報単位に対して所定の処理を行う

ことを特徴とする請求項 5 に記載の電子文書情報拡充方法。

【請求項 7】 上記外部データ取得工程は、抽出された所在情報に基づき、外部データを取得する際、外部データと、情報拡充対象の電子文書又は情報単位の内容とが類似していることを判定した上で取得することを特徴とする請求項 5 又は 6 に記載の電子文書情報拡充方法。

【請求項 8】 上記外部データ取得工程は、抽出された所在情報が示す場所の外部データと、その外部データが含む所在情報をたどって到達し得る外部データとを取得可能な外部データとしていることを特徴とする請求項 5 ～ 7 のいずれかに記載の電子文書情報拡充方法。

【請求項 9】 請求項 5 ～ 8 のいずれかに記載の電子文書情報拡充方法の各工程をコンピュータが処理し得るコードで記述したことを特徴とする電子文書情報拡充プログラム。

【請求項 10】 請求項 9 の電子文書情報拡充プログラムを記録していることを特徴とする記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、電子文書に対し、それが有しない要素の情報を拡充する電子文書情

報拡充装置に関し、例えば、電子メール文書を情報源として扱う情報マネジメントシステムに適用し得るものである。

【 0 0 0 2 】

【従来の技術】

近年、電子メールの文書中に、関連する情報の所在（例えばURL、URIなど；以下、この項の説明ではURLとする）を記述して送信することが日常的に行われるようになってきている。電子メール閲覧ソフトは、それに対応して、そのURLを選択するだけで、Webブラウザソフトが起動するように連携するなどさまざまな工夫がされてきている。ところが、電子メールを着信した時点では、まだURLで示される場所にある情報は取得しておらず、利用者が改めてその情報を取得する動作を行う必要がある。

【 0 0 0 3 】

【特許文献1】 特開 2 0 0 1 - 1 8 4 2 7 7 号公報

この問題点に着目して、特許文献1では、電子メールの中に参照すべき情報の所在がURLで示されていると、そのURLが示す場所にある情報（HTML文書など）を自動的に取得し、受信した電子メールと関連付けて保存しておくという方法が紹介されている。これにより、電子メールを受信したユーザは、ネットワークとの接続が切れた状態でも、電子メール文書中のURLを指定するだけで、既に取得済みのデータを表示装置によって閲覧することができる。

【 0 0 0 4 】

【発明が解決しようとする課題】

上記特許文献1の方法によると、電子メール文書内に含まれるURLのデータを、電子メールと関連付けて全て取得するため、実際には、電子メール文書の内容とは関係のない部分まで取得する恐れがある。したがって、全てのURLのデータを蓄積するのは、ネットワークとの接続が切れた状態でも、URLのデータを閲覧できる利点がある反面、蓄積効率が悪くなるという課題がある。

【 0 0 0 5 】

さらに、企業のURLが示されているときには、企業のトップページがリンクされていることが多く、このトップページのデータが蓄積されていると、電子メ

ール文書の内容に関連する情報を、トップページからのリンクをたどって探さなければならない。上記特許文献 1 の方法によると、指定した URL のページを取得して蓄積しているため、ネットワークの接続が切れた状態では、さらなるリンク先を探すことができないという課題が残されている。

【 0 0 0 6 】

そのため、関連する情報の所在を含む電子文書の情報を拡充できる電子文書情報拡充装置、方法及びプログラム、並びに、電子文書情報拡充プログラムを記録した記録媒体が望まれている。

【 0 0 0 7 】

【課題を解決するための手段】

第 1 の本発明の電子文書情報拡充装置は、電子文書に対し、それが有しない要素の情報を拡充するものであって、（１）入力された電子文書から、それに含まれているデータの所在情報を抽出する所在情報抽出手段と、（２）抽出された所在情報に基づき、補充し得る外部データを取得する外部データ取得手段と、（３）取得した外部データから、入力された上記電子文書で不足する要素の情報を拡充する情報補完手段とを備えることを特徴とする。

【 0 0 0 8 】

第 2 の本発明の電子文書情報拡充方法は、電子文書に対し、それが有しない要素の情報を拡充するものであって、（１）入力された電子文書から、それに含まれているデータの所在情報を抽出する所在情報抽出工程と、（２）抽出された所在情報に基づき、補充し得る外部データを取得する外部データ取得工程と、（３）取得した外部データから、入力された上記電子文書で不足する要素の情報を拡充する情報補完工程とを有することを特徴とする。

【 0 0 0 9 】

第 3 の本発明の電子文書情報拡充プログラムは、第 2 の本発明の電子文書情報拡充方法の各工程をコンピュータが処理し得るコードで記述したことを特徴とする。

【 0 0 1 0 】

第 4 の本発明の記録媒体は、第 3 の本発明の電子文書情報拡充プログラムを記

録していることを特徴とする。

【 0 0 1 1 】

【発明の実施の形態】

(A) 実施形態

以下、本発明による電子文書情報拡充装置、方法及びプログラム、並びに、電子文書情報拡充プログラムを記録した記録媒体の一実施形態を図面を参照しながら詳述する。

【 0 0 1 2 】

この実施形態は、電子メール文書に対して、URLで示された情報源にアクセスし、個々の情報と関連する内容を情報源から取得した上でキーワード抽出を行い、その結果を含めて構造化データを生成するものである。

【 0 0 1 3 】

(A-1) 実施形態の構成

図1は、この実施形態の電子文書情報拡充装置の機能的構成を示すブロック図である。

【 0 0 1 4 】

例えば、実施形態の電子文書情報拡充装置は、通信機能を有するパソコン等のユーザの情報処理装置（メールクライアント）に対し、CD-ROMやフロッピー（登録商標）ディスク等の記録媒体に記録されている電子文書情報拡充プログラム（例えば、電子メール閲覧ソフトの追加機能となっている）をインストールすることで実現されるが、機能的には、図1で表すことができる。また例えば、メールサーバに対し、CD-ROMやフロッピーディスク等の記録媒体に記録されている電子文書情報拡充プログラムをインストールすることで実現されるが、この場合にも、機能的には、図1で表すことができる。

【 0 0 1 5 】

この実施形態の電子文書情報拡充装置は、入力部100、情報解析部101、外部データ取得部102、情報補完部（情報補充部）103及び構造化データ生成部104を有する。

【 0 0 1 6 】

入力部 1 0 0 は、情報、及び、その情報に関連する情報の情報源を示した URL (URI、FTP、ファイル名等でも良いが、以下では、URL で説明する) を含む電子メール文書 (例えば、メールマガジン) の入力を行うものである。電子メール文書の入力は、この入力時に取り出したものであっても良く、過去に取り出して記憶しているものを読み出したものであっても良い。

【 0 0 1 7 】

情報解析部 1 0 1 は、入力された電子メール文書を、個々の情報単位に分割し、それぞれの情報単位から、情報源を示す URL を抽出するものである。例えば、電子メール文書がニュースのメールマガジンであれば、1 つの記事を 1 単位とする情報に分割する。そして、それぞれの情報単位の中に含まれる URL を抽出する。

【 0 0 1 8 】

外部データ取得部 1 0 2 は、情報解析部 1 0 1 で分割された情報単位毎に、そこに含まれている URL に基づいて、情報単位内に記載されている内容と類似する詳細なデータを URL など示される外部の情報源から取得するものである。外部データ取得部 1 0 2 は、取得するに値するデータの判断は、情報単位内に記載されている元の文章と、URL など示される情報源から取得したデータとの類似度を見ることで行う。

【 0 0 1 9 】

情報補完部 1 0 3 は、外部データ取得部 1 0 2 で取得したデータに対して、キーワードの抽出と主要箇所の抽出を行い、元の情報単位に追加する補充データを作成するものである。

【 0 0 2 0 】

構造化データ生成部 1 0 4 は、情報補完部 1 0 3 がそろえた補充すべきデータをまとめて、構造化データを生成するものである。

【 0 0 2 1 】

(A-2) 実施形態の動作

図 2 は、この実施形態の電子文書情報拡充装置の全体動作 (電子文書情報拡充方法) を示すフローチャートである。

【 0 0 2 2 】

ここでは、情報単位の例として、タイトル<TITLE>、概要<BODY>、キーワード<KEYWORD>、情報源の場所<URL>を、情報単位を構成する必須の内容とし、これらを全て含む構造化データの生成について説明する。また、キーワードは全ての場合に生成されることになるが、電子メール文書に対して分割処理を行った後、概要が欠けている場合について以降で詳細に説明する。

【 0 0 2 3 】

ステップ S 2 0 0 の入力処理では、入力部 1 0 0 が電子メール文書を入力する。

【 0 0 2 4 】

ステップ S 2 0 1 の情報単位抽出処理では、情報解析部 1 0 1 が、入力された電子メール文書の中に含まれている情報を、それぞれ関連する内容毎に切り分ける。例えば、電子メール文書が、図 3 に示されているような場合には、図 4 に示すような情報単位に切り分けられる。このとき、切り分ける方法は、セパレータと呼ばれる特殊な記号の連続や、空白行などを手がかりに、これらで挟まれた部分を情報単位としても良い。また、段落や、表題記号などを手がかりにして、次の段落や表題記号が出現するまでを 1 つの情報単位としても良い。

【 0 0 2 5 】

そして、切り分けられた情報単位の中に、その情報の詳細情報の所在を示す URL が記載されていれば抽出する。

【 0 0 2 6 】

この実施形態では、抽出した結果をタグでマーキングするという形で表現している。例えば、図 4 に示す情報単位に対しては、図 5 に示すように抽出し、表現する。例えば、情報単位の 1 行目はタイトルと認識している。また、1 つの情報単位の中に、URL が複数個存在する場合にも、同様に抽出する。但し、複数の出現については、URL の表現を区別するために、タグの中に id という属性を付与し出力順に番号を付与している。URL の発見方法については、http://で始まる文字列を検索するなど、一般的な方法を利用すれば良い。また、U

R L 抽出後の表現方法も、複数の URL を識別できるのであれば、上記の方法に限るものではない。

【 0 0 2 7 】

ステップ S 2 0 2 ～ S 2 0 7 の処理は、抽出された情報単位毎に実行される処理である。

【 0 0 2 8 】

ステップ S 2 0 2 のデータ取得処理（情報取得処理）では、外部データ取得部 1 0 2 が、ステップ S 2 0 1 で抽出した URL に基づいて、その URL が示す情報源等からデータを取得する。これは一般に、ネットワークを介して URL が示すサーバにアクセスし、該当する HTML 文書を取得するという処理になる。

【 0 0 2 9 】

ステップ S 2 0 3 の判定処理では、データ取得処理 S 2 0 2 で取得した URL が示す場所のデータと、その URL が含まれていた情報単位の内容とが一致するかを判定する。判定の方法としては、例えば、取得したデータと、情報単位の内容のそれぞれからキーワードを抽出し、互いのキーワードの一致度を計算し、一致度を閾値と比較する方法で行う。一致と判定した場合には、処理をステップ S 2 0 5 に移行し、不一致と判定した場合には、処理をステップ S 2 0 4 に移行する。

【 0 0 3 0 】

図 6 は、図 5 の 2 番目の情報単位に取得したデータを追加した様子を示したものであり、取得したデータに＜GET-DATA＞というタグを付けて表現したものである。

【 0 0 3 1 】

ここで、取得したデータは、一般的に HTML 文書と呼ばれる制御文字を含んだ文書である。そのため、予め取得したデータのハイパーリンク以外の制御文字を取り除く前処理を行った上で、判定処理を行っても良い。

【 0 0 3 2 】

また、取得したデータは、レイアウトなどによって記載内容を分類できるため、予め取得したデータの主要箇所を抽出する前処理を行った上で、抽出された主

要箇所との判定処理を行っても良い。

【 0 0 3 3 】

ステップ S 2 0 3 の判定処理において、不一致と判定された場合に移行する、ステップ S 2 0 4 の URL 変更処理では、先に取得したデータの中に含まれる全てのハイパーリンクを抽出して、第 1 階層の URL リストを作成した上で一時的に記憶しておき、各 URL に対してデータ取得処理 S 2 0 2 及び判定処理 S 2 0 3 を繰り返す。第 1 階層の URL リストに対しても全て不一致となった場合は、一時的に記憶されている第 1 階層の URL リストから取得できるデータに対して、再びハイパーリンクを抽出して第 2 階層の URL リストを作成して一時的に記憶しておき、各 URL に対してデータ取得処理 S 2 0 2 及び判定処理 S 2 0 3 を繰り返す。

【 0 0 3 4 】

例えば、情報単位に含まれている URL が企業のトップページのような場合には、トップページに含まれている全てのハイパーリンクを取り出して、そのリンク先の Web ページに移行して、情報単位に関連するかを判定し、第 1 階層の URL に係る Web ページも情報単位に関連しない場合には、その各 Web ページに含まれている全てのハイパーリンクを取り出し、情報単位に関連する Web ページの探索を行う。

【 0 0 3 5 】

ここで、探索をやめるまでの階層の深さは予め固定設定しておいても良く、ユーザが任意に設定できるようにしても良く、いずれにせよ、繰り返しの回数を制限できるものとする。

【 0 0 3 6 】

なお、抽出した情報単位に複数の URL が記載されている場合には、ある 1 個の URL についてデータ取得を行い、取得データが情報単位に関係しないと判定された場合には、次の URL でのデータ取得、判定を行い、一致するものが見つかるまで繰り返す。しかし、全ての URL での取得データが不一致のときには、ある 1 個の URL の取得データに対する上述した第 1 階層のリンクの処理を行い、これでも一致する取得データがなければ、他の URL の取得データに対する上

述した第1階層のリンクの処理を行い、以下、情報単位の内容に一致する取得データを得るまで（階層の制限はある）、同様な処理を繰り返す。上述とは異なり、各URLのそれぞれについて、取得データを得て、その中で一致度が最も高いものを選択するようにしても良い。

【0037】

また、ステップS201で抽出された情報単位にURLが含まれていない場合には、その情報単位に対するステップS202～S207の処理を省略しても良い。また、電子メール文書（例えばメールマガジン）を提供している企業の代表のURLや、新聞社等のURL等が情報単位に含まれているものとみなして（このようなURLはシステムが固定設定しても良く、ユーザが任意設定しても良い）、処理を行うようにしても良い。この場合の探索階層の深さは、情報単位にURLが含まれている場合と同様でも良いが、より深くしても構わない。

【0038】

情報単位の内容に関連する取得データが得られた場合には、処理はステップS205に移行する。情報単位の内容に関連する取得データが得られなかった場合には、ステップS205に移行しないで、次の情報単位の処理に移行しても良く、ステップS205に移行し、情報単位に係る処理だけを行うようにしても良い（取得データに対する処理は実行されない）。

【0039】

ステップS205のキーワード抽出処理は、情報補完部103の処理の一つで、情報単位の中に含まれる内容と、取得したデータに対して、キーワードとして扱う文字列の抽出を行う。なお、ステップS203の判定処理でキーワード抽出を行っているならばそれを利用するようにしても良い。キーワードの抽出方法については、特に限定されるものではなく、既存の方法を用いて良い。但し、情報単位の中に含まれるキーワードであるのか、取得したデータの中に含まれるキーワードであるのかは、情報単位に対する検索を行う際に検索対象を選択可能にするため、区別して情報単位を管理する。

【0040】

例えば、図7に示すように、情報単位の中から抽出したキーワードと、取得デ

ータから抽出したキーワードとにキーワードであることを表すタグを付けると共に、キーワードタグの属性値を利用して、どちらから抽出したキーワードであるかを表して情報単位の中に表現しておく。例えば、キーワードが情報単位の中に含まれるものであれば、属性値をT（タイトル部分）やD（概要部分）とし、取得したデータからであれば属性値をGとし、複数の部分に含まれているものであれば、その記号を列挙する。

【 0 0 4 1 】

ステップS 2 0 6の主要部抽出処理は、情報補完部 1 0 3の処理の一つで、取得したデータの中から、主要部のみを抽出する。なお、主要部抽出方法としても、既存の方法を適用することができ、既存の要約作成方法を適用することができる。ここでの主要部とは、取得したデータの中から、情報単位が示す内容と類似する、又は、情報単位の示す内容の詳細に相当する部分をいう。主要部として抽出する文字数を制限しなければ、取得したデータを全て主要部として扱っても良いが、具体的にはある文字数で制限し、その文字数に収まるように、取得したデータから主要部を抽出する。

【 0 0 4 2 】

例えば、図 8 に示すように、＜GET-DATA＞というタグで囲むことにより表現されている取得したデータから、主要部を抽出し、＜BODY＞というタグで囲んで、情報単位の中に表現する。このとき、この主要部が取得したデータから得られたものであることを示す情報として、属性値にGという記号をつけておく。情報単位内にもともと主要部（又は要約）含まれていた場合には、属性値にOという記号をつけておく。

【 0 0 4 3 】

ステップS 2 0 7の構造化データ作成処理は、構造化データ生成部 1 0 4の処理で、情報単位の内容と、キーワード抽出処理（S 2 0 4）の結果と、主要部抽出処理（S 2 0 5）の結果を組み合わせ、構造化データを生成する。例えば、図 9 に示すように、タグ付けした形で、構造化データを生成する。このとき、取得したデータについては、不要なものも含まれているため、主要部を抽出した後は削除することで、蓄積効率を上げる。なお、取得データを残しても良いことは

勿論である。

【 0 0 4 4 】

ステップ S 2 0 8 の判定処理では、情報単位抽出処理（S 2 0 1）において、複数の情報単位が抽出された場合に、処理されていない情報単位があるか否かを判定する。処理されていない情報単位があれば、ステップ S 2 0 2 に処理を移行する。

【 0 0 4 5 】

全ての情報単位に対して処理されていれば、生成した全ての構造化データを出力する。出力は、表示出力でも、印刷出力でも、送信出力でも良く、また、後の表示出力や印刷出力などのための記憶処理でも良い。また、生成した全ての構造化データを出力するのではなく、ユーザが予め指定したキーワードを含む構造化データを出力するようにしても良い。

【 0 0 4 6 】

（A-3）実施形態の効果

上記実施形態によれば、メールサーバ若しくはメールクライアントの 1 機能として動作させることにより、電子メール文書の中に、URL で示されている部分があった場合、URL で示された部分から、電子メール文書の内容に該当するデータが読み込まれた状態で出力できるため、ユーザは、改めて URL の指定や、URL の情報を取得する必要がなく、十分な必要な情報を得ることができる。特に、メールサーバに拡充機能を持たせた場合には、ユーザは、メール受信時に何らの操作も行わずに、十分な情報を得ることができる。

【 0 0 4 7 】

また、取得したデータ全てを蓄積するのではなく、電子メール文書の内容に相当するデータの中から、さらに主要部のみを抽出して蓄積するため、蓄積効率が良い。

【 0 0 4 8 】

さらに、電子メールの受信だけで、URL の情報も同時に取得できているので、電子メールの閲覧ソフトウェアだけで、必要な URL 情報も閲覧することができる。

【 0 0 4 9 】

さらにまた、タイトルとURLだけからなるような情報に対して、URLが示すサーバから取得したデータに対してキーワード抽出を行った上で、構造化データを作成するため、本構造化データをデータベース等に蓄積した上でキーワード検索をする際、検索精度が、タイトルだけに対して検索を行った場合に比べて、格段に向上する。

【 0 0 5 0 】

(B) 他の実施形態

上記実施形態の電子文書情報拡充装置からの最終出力の形式は、必要に応じて、電子メール文書の形態や、Webブラウザで閲覧可能な形態に変換しても良い。そして、これらを電子メールとしてユーザに送信しても良い。すなわち、拡充後の情報単位は、構造化データになっていなくても良い。

【 0 0 5 1 】

また、情報単位の内容とURLが示すサーバから取得したデータとの類似度（一致度）を判定する際、予め指定した深さの階層までの全てのリンク先のデータを取得して類似度を計算した上で、最も類似度の高いデータを採用するようにしても良い。

【 0 0 5 2 】

さらに、ステップS205のキーワード抽出処理とステップS206の主要部抽出処理の順序を入れ替えても良い。その場合は、主要部抽出処理の結果に対して、キーワード抽出処理を行うことになる。

【 0 0 5 3 】

さらにまた、入力される電子メール文書が、複数の情報を含んでいない場合でも良く、このような電子メール文書を専用に扱う装置であれば、分割処理手段を備えていなくても良い。また、本発明の電子文書は、電子メール文書に限定されず、入力文書そのものがWebページ等であっても良く、その場合、タグを外して上記と同様な処理を行っても良く、上記で用いるタグはそのまま残して処理しても良い。電子文書は、コンテンツとして提供されるようなものであっても良い。また、既に、情報単位に分割されたものが入力され、各情報単位に対して情

報の拡充を行うものであっても良い。

【 0 0 5 4 】

上記実施形態では、URLが情報の所在を表すものであったが、URIやFTPやファイル名等であっても良い。

【 0 0 5 5 】

また、上記実施形態では、取得データの詳細を最終的には除去するものを示したが、除去するか否かをユーザに予め設定させるようにしても良い。すなわち、拡充情報は、主要部及びキーワードに限定されず、取得データの詳細情報を含めても良く、キーワードだけを拡充するものであっても良く、ユーザに任意に設定させるものであっても良い。

【 0 0 5 6 】

さらに、上記実施形態では、情報を拡充する場合を示したが、情報を置換するようにしても良い。例えば、情報単位に概要が含まれていても、取り出した取得データでの概要の方が詳しい場合には（例えば、文字数や文章数で判断する）、情報単位に含まれている概要を、取得データの概要に置換するようにしても良い。

【 0 0 5 7 】

さらにまた、上記実施形態では、情報を拡充する場合を示したが、拡充の際に、拡充情報や当初情報を翻訳するようにしても良い。例えば、取り出した取得データが、外国語（当初情報から見た場合の外国語や利用者の指定言語と異なる外国語）であった場合に、利用者が理解できる言語等に翻訳した上で、拡充するようにしても良い。また例えば、両言語の情報を並記するようにしても良い。

【 0 0 5 8 】

特許請求の範囲の「拡充」の用語には、このような置換や翻訳による情報量の拡充も含むものとする。

【 0 0 5 9 】

【発明の効果】

以上のように、本発明によれば、関連する情報の所在を含む電子文書の情報を拡充できる電子文書情報拡充装置、方法及びプログラム、並びに、電子文書情報

拡充プログラムを記録した記録媒体を提供できる。

【図面の簡単な説明】

【図 1】

実施形態の電子文書情報拡充装置（電子メール文書情報拡充装置）の機能的構成を示すブロック図である。

【図 2】

実施形態の電子メール文書情報拡充装置の全体動作を示すフローチャートである。

【図 3】

電子メール文書の一例を示す説明図である。

【図 4】

実施形態の情報解析部による図 3 の文書に対する情報単位抽出処理の結果例を示す説明図である。

【図 5】

実施形態の情報解析部による抽出情報単位に対する URL の抽出結果例を示す説明図である。

【図 6】

実施形態の外部データ取得部の取得結果例を示す説明図である。

【図 7】

実施形態のキーワード抽出処理の処理結果例を示す説明図である。

【図 8】

実施形態の主要部抽出処理の処理結果例を示す説明図である。

【図 9】

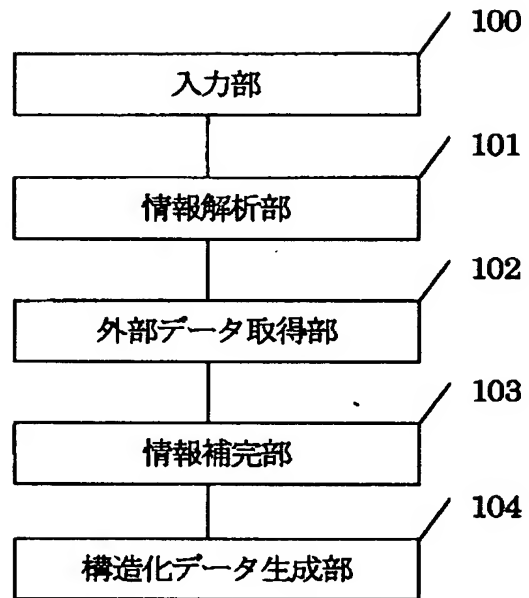
実施形態の構造化データ生成処理による構造化データ例を示す説明図である。

【符号の説明】

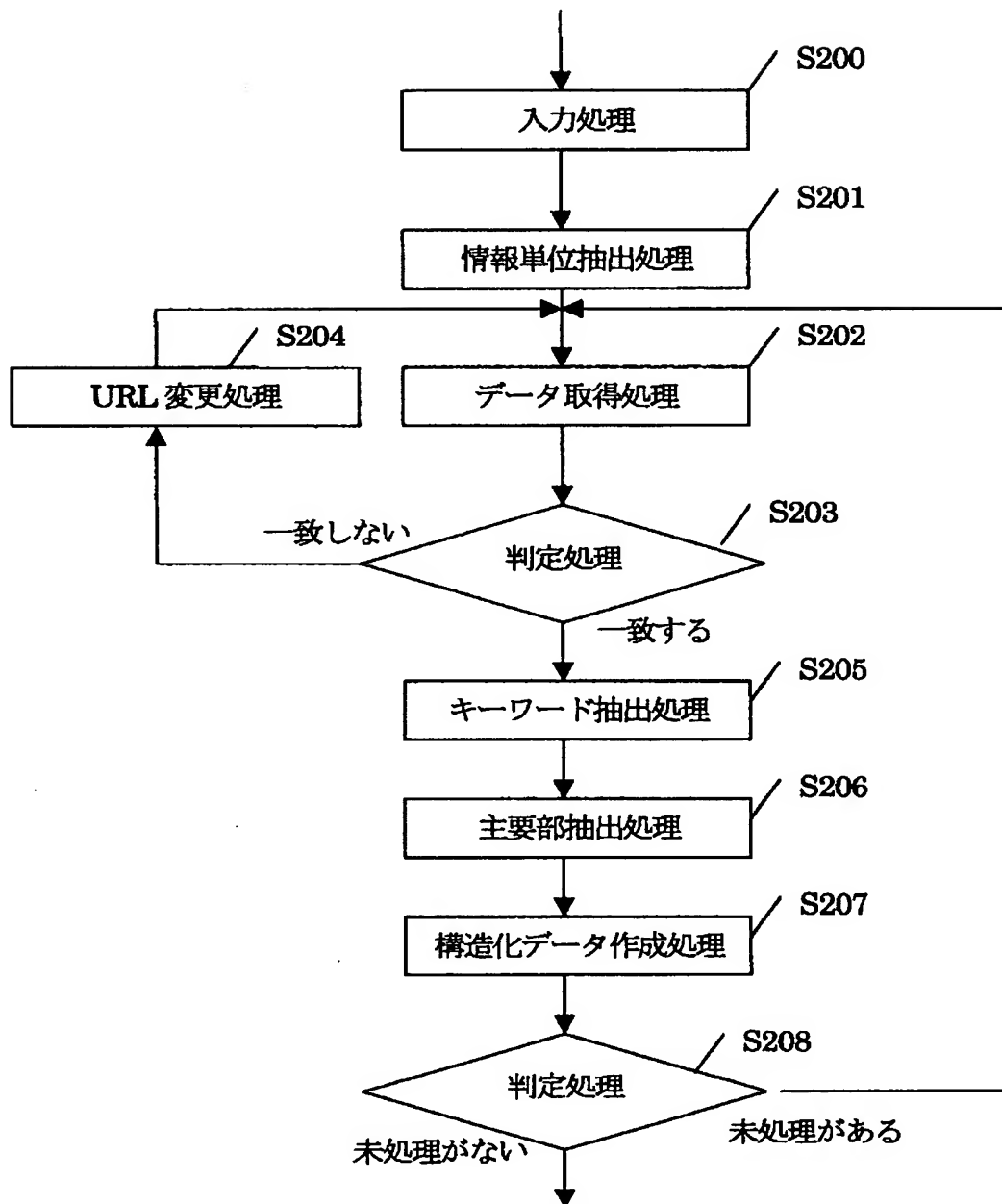
1 0 0 …入力部、1 0 1 …情報解析部、1 0 2 …外部データ取得部、1 0 3 …情報補完部（情報補充部）、1 0 4 …構造化データ生成部。

【書類名】 図面

【図 1】



【図 2】



【図 3】

<p>.....</p> <p>〇〇会社、△△を発表 〇月×日、〇〇会社は、新製品△△を発表した。価格は、2500 円～10000 円。全国の家電量販店等で〇月×日より発売を開始する。 http://www.xxx.zz/message1.htm</p> <p>.....</p> <p>××銀行と△△銀行が合併 http://www.yyy.zz/message2.htm</p> <p>.....</p> <p>情報系新会社△△研究所が誕生 http://www.zzz.zz/message3.htm http://www.aaa.zz/release1.htm</p> <p>.....</p> <p>：</p> <p>：</p> <p>：</p>
--

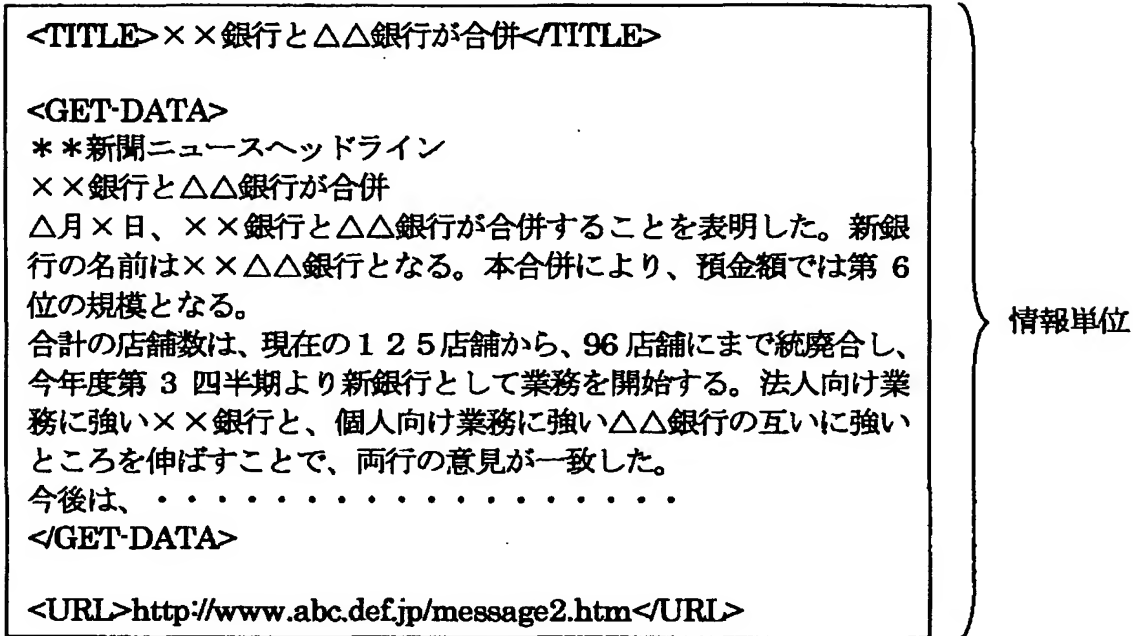
【図 4】

<p>〇〇会社、△△を発表 〇月×日、〇〇会社は、新製品△△を発表した。価格は、2500 円 ～10000 円。全国の家電量販店等で〇月×日より発売を開始する。 http://www.abc.def.jp/message1.htm</p>	情報単位
<p>××銀行と△△銀行が合併 http://www.abc.def.jp/message2.htm</p>	情報単位
<p>情報系新会社△△研究所が誕生 http://www.abc.def.jp/message3.html http://www.aaa.zz/release1.htm</p>	情報単位

【図 5】

<p><TITLE>〇〇会社、△△を発表</TITLE> <BODY> 〇月×日、〇〇会社は、新製品△△を発表した。価格は、2500 円 ～10000 円。全国の家電量販店等で〇月×日より発売を開始する。 </BODY> <URL id=1>http://www.abc.def.jp/message1.htm</URL></p>
<p><TITLE>××銀行と△△銀行が合併</TITLE> <URL id=1>http://www.abc.def.jp/message2.htm</URL></p>
<p><TITLE>情報系新会社△△研究所が誕生</TITLE> <URL id=1>http://www.abc.def.jp/message3.html</URL> <URL id=2>http://www.aaa.zz/release1.htm</URL></p>

【図 6】



【図 7】

<TITLE>××銀行と△△銀行が合併</TITLE>

<GET-DATA>

**新聞ニュースヘッドライン

××銀行と△△銀行が合併

△月×日、××銀行と△△銀行が合併することを表明した。新銀行の名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。

合計の店舗数は、現在の 125 店舗から、96 店舗にまで統廃合し、今年度第 3 四半期より新銀行として業務を開始する。法人向け業務に強い××銀行と、個人向け業務に強い△△銀行の互いに強いところを伸ばすことで、両行の意見が一致した。

今後は、.....

</GET-DATA>

<KEYWORDS>

<KEYWORD att=TG>××銀行</KEYWORD>

<KEYWORD att=TG>△△銀行</KEYWORD>

<KEYWORD att=G>××△△銀行</KEYWORD>

<KEYWORD att=G>第 6 位</KEYWORD>

<KEYWORD att=G>125 店舗 </KEYWORD>

<KEYWORD att=G>96 店舗 </KEYWORD>

⋮

<KEYWORDS>

<URL>http://www.abc.def.jp/message2.htm</URL>

情報単位

【図 8】

```

<TITLE>××銀行と△△銀行が合併</TITLE>
<BODY att=G>
△月×日、××銀行と△△銀行が合併することを表明した。新銀行の名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。
</BODY>
<GET-DATA>
**新聞ニュースヘッドライン
××銀行と△△銀行が合併
△月×日、××銀行と△△銀行が合併することを表明した。新銀行の名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。
合計の店舗数は、現在の 125 店舗から、96 店舗にまで統廃合し、今年度第 3 四半期より新銀行として業務を開始する。法人向け業務に強い××銀行と、個人向け業務に強い△△銀行の互いに強いところを伸ばすことで、両行の意見が一致した。
今後は、・・・・・・・・・・・・・・・・
</GET-DATA>
<KEYWORDS>
  <KEYWORD att=TG>××銀行</KEYWORD>
  <KEYWORD att=TG>△△銀行</KEYWORD>
  <KEYWORD att=G>××△△銀行</KEYWORD>
  <KEYWORD att=G>第 6 位</KEYWORD>
  <KEYWORD att=G>125 店舗 </KEYWORD>
  <KEYWORD att=G>96 店舗 </KEYWORD>
  :
  :
</KEYWORDS>
<URL>http://www.abc.def.jp/message2.htm</URL>

```

情報単位

【図 9】

```
<INFO>
<TITLE>××銀行と△△銀行が合併</TITLE>
<BODY att=G>
△月×日、××銀行と△△銀行が合併することを表明した。新銀行の
名前は××△△銀行となる。本合併により、預金額では第 6 位の規模となる。
</BODY>
<KEYWORDS>
  <KEYWORD att=TG>××銀行</KEYWORD>
  <KEYWORD att=TG>△△銀行</KEYWORD>
  <KEYWORD att=G>××△△銀行</KEYWORD>
  <KEYWORD att=G>第 6 位</KEYWORD>
  <KEYWORD att=G>125 店舗 </KEYWORD>
  <KEYWORD att=G>96 店舗 </KEYWORD>
  :
  :
</KEYWORDS>
<URL>http://www.abc.def.jp/message2.htm</URL>
</INFO>
```

情報単位

【書類名】 要約書

【要約】

【課題】 関連するデータの所在情報を含む電子文書の情報を拡充する。

【解決手段】 本発明では、入力された電子文書から、それに含まれているデータの所在情報を抽出し、抽出した所在情報に基づき、補充し得る外部データを取得し、取得した外部データから、入力された電子文書で不足する要素の情報を拡充する。

【選択図】 図 1

出 願 人 履 歴 情 報

識別番号 [0 0 0 0 0 0 2 9 5]

1. 変更年月日	1 9 9 0 年 8 月 2 2 日
[変更理由]	新規登録
住 所	東京都港区虎ノ門1丁目7番12号
氏 名	沖電気工業株式会社